

MENU **SEARCH** **INDEX** **DETAIL** **JAPANESE** **BACK**

2 / 4

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-230021

(43)Date of publication of application : 16.08.2002

(51)Int.Cl. G06F 17/30

(21)Application number : 2001-021796 (71)Applicant : CANON INC

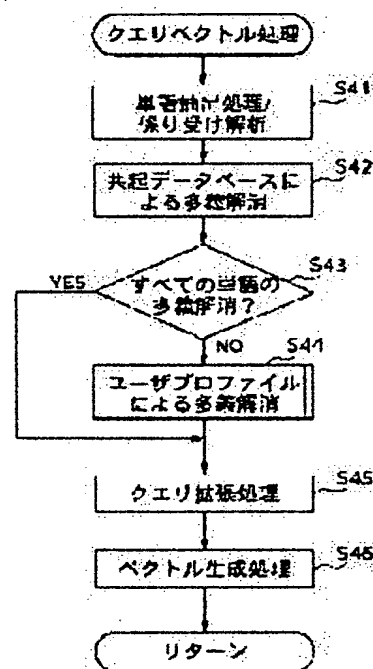
(22)Date of filing : 30.01.2001 (72)Inventor : TOSHIMA EIICHIRO

(54) INFORMATION RETRIEVAL DEVICE AND METHOD, AND STORAGE MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To easily and quickly perform information retrieval intended by a user.

SOLUTION: A word is extracted from a retrieval query designated by a user, and modification analysis is operated (S41), and polysemy elimination is tried based on the result of the modification analysis and a co-occurrence data base (S42), and when the polysemy is eliminated, the processing is immediately moved to a S45, and in the other case, polysemy elimination based on user profile information is operated, and then the processing is moved to the S45. Then, extension processing such as near-synonym development is operated by using a query extension dictionary (S45), and the document vector generation processing of the retrieval query is operated (S46).



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-230021
(P2002-230021A)

(43) 公開日 平成14年8月16日 (2002.8.16)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/30	3 3 0	G 0 6 F 17/30	3 3 0 C 5 B 0 7 5
	1 7 0		1 7 0 A
	3 2 0		3 2 0 D
	3 4 0		3 4 0 A
	3 5 0		3 5 0 C

審査請求 未請求 請求項の数27 O L (全 14 頁)

(21) 出願番号 特願2001-21796(P2001-21796)

(22) 出願日 平成13年1月30日 (2001.1.30)

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72) 発明者 戸島 英一朗

東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(74) 代理人 100081880

弁理士 渡部 敏彦

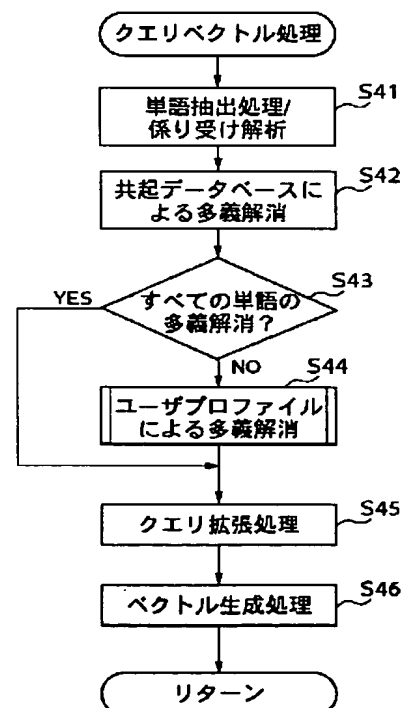
Fターム(参考) 5B075 ND03 NK35 PP24 PQ02 PQ36
PR06 PR08 QM08 QP03

(54) 【発明の名称】 情報検索装置及び情報検索方法並びに記憶媒体

(57) 【要約】

【課題】 ユーザの意図に即した情報検索を容易且つ迅速に行うことができるようにした。

【解決手段】 ユーザ指定の検索クエリから単語を抽出し、係り受け解析を行った後 (S 4 1)、係り受け解析の結果と共起データベースとに基づいて多義解消を試み (S 4 2)、多義性が解消された場合は直ちに S 4 5 に進む一方、解消されなかった場合はユーザプロフィール情報に基づく多義解消を行い、その後、S 4 5 に進む。そして、クエリ拡張辞書を使用し類義語展開等の拡張処理を行った後 (S 4 5)、検索クエリの文書ベクトル生成処理を行う (S 4 6)。



【特許請求の範囲】

【請求項 1】 検索条件を入力する検索条件入力手段と、該検索条件入力手段により入力された検索条件から単語を抽出する形態素解析手段と、複数の単語間の共起関係を語義と関連付けて記憶する共起関係記憶手段と、該共起関係記憶手段に記憶された共起情報と前記形態素解析手段の解析結果とに基づいて前記検索条件の共起関係を抽出し多義性を解消する第 1 の多義性解消手段と、該第 1 の多義性解消手段により多義性の解消された語義に基づいて情報検索を行う情報検索手段とを有することを特徴とする情報検索装置。

【請求項 2】 前記第 1 の多義性解消手段により多義性の解消された語義に対し拡張処理を行う語義拡張手段を備えていることを特徴とする請求項 1 記載の情報検索装置。

【請求項 3】 前記語義拡張手段は、少なくとも語義を類義語展開する類義語展開手段と、関連語に展開する関連語展開手段と、これらを組み合わせた組み合わせ展開手段とを含んでいることを特徴とする請求項 2 記載の情報検索装置。

【請求項 4】 ユーザの嗜好を表現するユーザプロフィール情報が記憶されたユーザプロフィール記憶手段と、前記ユーザプロフィール記憶手段に記憶されたユーザプロフィール情報と前記形態素解析手段の解析結果とに基づいて前記検索条件の多義性を解消する第 2 の多義性解消手段とを備え、前記第 1 の多義性解消手段は前記第 2 の多義性解消手段より優先することを特徴とする請求項 1 乃至請求項 3 のいずれかに記載の情報検索装置。

【請求項 5】 前記第 2 の多義性解消手段により多義性の解消された語義に対し拡張処理を行う語義拡張手段を備えていることを特徴とする請求項 4 記載の情報検索装置。

【請求項 6】 前記語義拡張手段は、少なくとも語義を類義語展開する類義語展開手段と、関連語に展開する関連語展開手段と、これらを組み合わせた組み合わせ展開手段とを含んでいることを特徴とする請求項 5 記載の情報検索装置。

【請求項 7】 前記第 1 の多義性解消手段により多義性の解消された語義に基づいて文書ベクトルを生成する文書ベクトル生成手段と、該文書ベクトル生成手段により生成された文書ベクトルをデータベースとして登録する登録手段とを有していることを特徴とする請求項 1 乃至請求項 6 のいずれかに記載の情報検索装置。

【請求項 8】 前記第 1 の多義性解消手段により多義性の解消された語義に基づいて文書ベクトルを生成する文書ベクトル生成手段と、ユーザの嗜好を表現するユーザプロフィール情報が記憶されたユーザプロフィール記憶手段の記憶内容を前記文書ベクトル生成手段により生成された文書ベクトルに基づいて更新するユーザプロフィール情報更新手段とを有していることを特徴とする請求

項 1 乃至請求項 7 のいずれかに記載の情報検索装置。

【請求項 9】 検索条件を入力する検索条件入力手段と、該検索条件入力手段により入力された検索条件から単語を抽出する形態素解析手段と、ユーザの嗜好を表現するユーザプロフィール情報が記憶されたユーザプロフィール記憶手段と、該ユーザプロフィール情報と前記形態素解析手段の解析結果とに基づいて前記検索条件の多義性を解消する第 2 の多義性解消手段と、該第 2 の多義性解消手段により多義性の解消された語義に基づいて情報検索を行う情報検索手段とを有することを特徴とする情報検索装置。

【請求項 10】 検索条件を入力する検索条件入力ステップと、該検索条件入力ステップで入力された検索条件から単語を抽出する形態素解析ステップと、複数の単語間の共起関係を語義と関連付けて記憶された共起情報と前記形態素解析ステップでの解析結果とに基づいて前記検索条件の共起関係を抽出し多義性を解消する第 1 の多義性解消ステップと、該第 1 の多義性解消ステップで多義性の解消された語義に基づいて情報検索を行う情報検索ステップとを含むことを特徴とする情報検索方法。

【請求項 11】 前記第 1 の多義性解消ステップで多義性の解消された語義に対し拡張処理を行う語義拡張ステップを含んでいることを特徴とする請求項 10 記載の情報検索方法。

【請求項 12】 前記語義拡張ステップは、少なくとも語義を類義語展開する類義語展開ステップと、関連語に展開する関連語展開ステップと、これらを組み合わせた組み合わせ展開ステップとを含んでいることを特徴とする請求項 11 記載の情報検索方法。

【請求項 13】 ユーザの嗜好を表現するユーザプロフィール情報と前記形態素解析ステップの解析結果とに基づいて前記検索条件の多義性を解消する第 2 の多義性解消ステップを備え、前記第 1 の多義性解消ステップは前記第 2 の多義性解消ステップより優先することを特徴とする請求項 10 乃至請求項 12 のいずれかに記載の情報検索方法。

【請求項 14】 前記第 2 の多義性解消ステップで多義性の解消された語義に対し拡張処理を行う語義拡張ステップを含んでいることを特徴とする請求項 13 記載の情報検索方法。

【請求項 15】 前記語義拡張ステップは、少なくとも語義を類義語展開する類義語展開ステップと、関連語に展開する関連語展開ステップと、これらを組み合わせた組み合わせ展開ステップとを含んでいることを特徴とする請求項 14 記載の情報検索装置。

【請求項 16】 前記第 1 の多義性解消ステップで多義性の解消された語義に基づいて文書ベクトルを生成する文書ベクトル生成ステップと、該生成された文書ベクトルをデータベースとして登録手段に登録する登録ステップとを含んでいることを特徴とする請求項 10 乃至請求

項 15 のいずれかに記載の情報検索方法。

【請求項 17】 前記第 1 の多義性解消ステップで多義性の解消された語義に基づいて文書ベクトルを生成する文書ベクトル生成ステップと、ユーザの嗜好を表現するユーザプロフィール情報が記憶されたユーザプロフィール記憶手段の記憶内容を前記生成された文書ベクトルに基づいて更新するユーザプロフィール情報更新ステップとを含んでいることを特徴とする請求項 10 乃至請求項 16 のいずれかに記載の情報検索方法。

【請求項 18】 検索条件を入力する検索条件入力ステップと、該検索条件入力手段により入力された検索条件から単語を抽出する形態素解析ステップと、ユーザの嗜好を表現するユーザプロフィール情報が記憶されたユーザプロフィール情報と前記形態素解析ステップの解析結果とに基づいて前記検索条件の多義性を解消する第 2 の多義性解消ステップと、該第 2 の多義性解消ステップにより多義性の解消された語義に基づいて情報検索を行う情報検索ステップとを含むことを特徴とする情報検索方法。

【請求項 19】 入力装置から入力された検索条件に基づいて情報検索を行う情報検索手順が記憶されたコンピュータ読み取り可能な記憶媒体であって、前記入力された検索条件から単語を抽出する形態素解析手順と、複数の単語間の共起関係を語義と関連付けて記憶された共起情報と前記形態素解析手順の解析結果とに基づいて前記検索条件の共起関係を抽出し多義性を解消する第 1 の多義性解消手順と、該第 1 の多義性解消手順により多義性の解消された語義に基づいて情報検索を行う情報検索手順とが記憶されていることを特徴とするコンピュータ読み取り可能な記憶媒体。

【請求項 20】 前記第 1 の多義性解消手順で多義性の解消された語義に対し拡張処理を行う語義拡張手順が記憶されていることを特徴とする請求項 19 記載の記憶媒体。

【請求項 21】 前記第 1 の多義性解消手順より優先される、ユーザの嗜好を表現するユーザプロフィール情報と前記係り受け解析ステップの解析結果とに基づいて前記検索条件の多義性を解消する第 2 の多義性解消手順が記憶されていることを特徴とする請求項 19 又は請求項 20 記載の記憶媒体。

【請求項 22】 前記第 2 の多義性解消手順で多義性の解消された語義に対し拡張処理を行う語義拡張手順が記憶されていることを特徴とする請求項 21 記載の記憶媒体。

【請求項 23】 前記第 1 の多義性解消手順で多義性の解消された語義に基づいて文書ベクトルを生成する文書ベクトル生成手順と、該生成された文書ベクトルをデータベースとして登録手段に登録する登録手順とが記憶されていることを特徴とする請求項 19 乃至請求項 22 のいずれかに記載の記憶媒体。

【請求項 24】 前記第 1 の多義性解消手順で多義性の解消された語義に基づいて文書ベクトルを生成する文書ベクトル生成手順と、ユーザの嗜好を表現するユーザプロフィール情報が記憶されたユーザプロフィール記憶手段の記憶内容を前記生成された文書ベクトルに基づいて更新するユーザプロフィール情報更新ステップとが記憶されていることを特徴とする請求項 19 乃至請求項 23 のいずれかに記載の記憶媒体。

【請求項 25】 語義ベースで単語の意味ベクトルが格納された単語ベクトル辞書と、検索対象となる文書データベースと、ユーザの嗜好を記録したユーザプロフィールと、係り受け単語間の共起関係が格納された共起データベースとからなるデータ構造を有していることを特徴とするコンピュータ読み取り可能な記憶媒体。

【請求項 26】 検索条件を拡張処理するための拡張辞書が格納されていることを特徴とする請求項 25 記載の記憶媒体。

【請求項 27】 入力装置から入力された検索条件に基づいて情報検索を行う情報検索手順が記憶されたコンピュータ読み取り可能な記憶媒体であって、前記入力された検索条件から単語を抽出する形態素解析手順と、ユーザの嗜好を表現するユーザプロフィール情報が記憶されたユーザプロフィール情報と前記形態素解析手順の解析結果とに基づいて前記検索条件の多義性を解消する第 2 の多義性解消手順と、該第 2 の多義性解消手順により多義性の解消された語義に基づいて情報検索を行う情報検索手順とが記憶されていることを特徴とするコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は情報検索装置及び情報検索方法並びに記憶媒体に関し、より詳しくは、入力された検索文や検索キーワード等の検索条件（以下、「クエリ」という）に従って情報検索を行う情報検索装置、及び該情報検索装置を使用した情報検索方法、並びに情報検索の検索手順や情報検索を行うためのデータ構造が記憶された記憶媒体に関する。

【0002】

【従来の技術】近年におけるコンピュータや通信網（ネットワーク）の発達に伴い、大量の電子化された文書のデータベースへの蓄積が進展してきており、それに伴って電子化された大量のデータベースから所望の文書データを検索する情報検索の需要が高まってきている。

【0003】この種の情報検索は、従来、キーワード検索、全文検索のようなクエリとの表記の一致を前提とした検索方法が主流であったが、最近では特定のクエリや文書に類似している類似文書の検索を行う手法が提案されている。

【0004】そして、このような類似文書の検索手法として、文書を n 次元のベクトル空間上の点にマッピング

し、それらの間の距離の大小により文書同士の類似性、又はクエリと文書との類似性を算出するベクトル空間モデル方式が既に知られている（例えば、「熊本、島田、加藤：概念ベースの情報検索への適用、信学技報、Vol. A198-63, pp. 9-16, 1999」）。

【0005】しかしながら、このようなベクトル空間モデル方式では、特に短い文章をクエリとして類似文書の検索を行った場合、ユーザの所望しない文書が検索されることも多い。

【0006】すなわち、クエリに使用される文字列は多義性を有するため、前記クエリがユーザの所望する意味に解釈されなかったり、或いはクエリに使用される文字列の語義の解釈が不完全であるためにクエリの意味が十分に補完されず、このため、上述したベクトル空間モデル方式では検索結果がユーザの意図通りにならない場合が多い。

【0007】そこで、斯かるユーザの意図しない文書の検索を極力回避する方策として、クエリを類義語で拡張する際に、単語間で共起性の低い類義語を展開対象から除外する技術が提案されている（例えば、特開平11-45274号公報；以下「第1の従来技術」という）。

【0008】該第1の従来技術では、前記共起性の低い類義語を文書検索の検索対象から取り除くことにより、ユーザの所望する文書とは無関係な類義語の出現している文書を検索対象から除外している。

【0009】また、その他の従来技術としては、各ユーザの文書ごとのアクセス状況を記録して該アクセス状況に基づいた嗜好ベクトルを作成し、クエリのベクトルを嗜好ベクトルに近付くようにシフトさせた技術も提案されている（特開平11-53394号公報；以下、「第2の従来技術」という）。

【0010】該第2の従来技術では、クエリのベクトルを嗜好ベクトルに近づくようにシフトさせることにより、検索結果がユーザの嗜好に近付くようにし、これによりユーザの嗜好を反映した文書の検索を可能にしている。

【0011】

【発明が解決しようとしている課題】しかしながら、上記第1の従来技術は、クエリの意味を拡張するために類義語展開する際に共起性の低い類義語を除外しているのみであるため、クエリ中の各単語の語義が特定されず、したがって各単語の多義性は依然として解消されず、ユーザの所望しない文書が検索結果として表示されることも多いという問題点があった。

【0012】すなわち、第1の従来技術では、例えば、ユーザがクエリとして「フォームの種類にどんなものがあるか教えてくれ」という文字列を入力した場合、帳票に関するドキュメントや野球の投球フォームに関するドキュメントを含め多くのドキュメントが検索され、クエリが多義性を保持した状態で表示出力される。

【0013】しかしながら、金融の帳票設計を業務とするユーザには、帳票の種類に関するドキュメントが必要であって野球の投球姿勢や水泳の泳法等に関するドキュメントは通常は必要としない。一方、スポーツを趣味とするユーザにとっては、野球の投球姿勢や水泳の泳法に関するドキュメントを所望する場合が多く、帳票の種類に関するドキュメントは通常は所望しない場合が多い。

【0014】すなわち、上記第1の従来技術では、共起性の低い類義語は排除されるものの、クエリの有する多義性を保持した状態で検索されるため、ユーザの意図した文書以外に多数の文書が表示出力され、このため所望の検索結果を容易且つ迅速に得ることができないという問題点があった。

【0015】また、第2の従来技術は、クエリ全体の意味をユーザの過去の嗜好、すなわち嗜好履歴の方向に強制的にシフトしているため、例えば、帳票の設計者が野球の投手の投球フォームを調べたいときには、システム上での語義解釈に誤解が生じないように、単なる「フォームの種類」ではなく、「投球フォームの種類」という文字列の入力を考慮する必要が生じ、このため使い勝手が悪くなるという問題点があった。

【0016】しかも、該第2の従来技術では、クエリが一律に嗜好ベクトルに近付くため、たとえ、「投球フォームの種類」の文字列を入力した場合であっても、帳票の設計者に対しては帳票に関するドキュメントが検索される虞もあり、その結果ユーザの所望しない検索結果が表示される場合があるという問題点があった。

【0017】このように上記第1及び第2の従来技術では、クエリの表記内容やユーザの嗜好に基づいて検索処理されているに過ぎず、クエリの有する多義性が考慮されず、またクエリの意味内容の補完も十分に行われていないため、ユーザの所望しない検索結果が得られることも多いという問題点があった。

【0018】本発明はこのような問題点に鑑みなされたものであって、ユーザの意図に即した情報検索を容易且つ迅速に行うことのできる情報検索装置及び情報検索方法並びに記憶媒体を提供することを目的とする。

【0019】

【課題を解決するための手段】上記目的を達成するために本発明に係る情報検索装置は、検索条件を入力する検索条件入力手段と、該検索条件入力手段により入力された検索条件から単語を抽出する形態素解析手段と、複数の単語間の共起関係を語義と関連付けて記憶する共起関係記憶手段と、該共起関係記憶手段に記憶された共起情報と前記形態素解析手段の解析結果とに基づいて前記検索条件の共起関係を抽出し多義性を解消する第1の多義性解消手段と、該第1の多義性解消手段により多義性の解消された語義に基づいて情報検索を行う情報検索手段とを有することを特徴とし、ユーザの嗜好を表現するユーザプロファイル情報が記憶されたユーザプロファイル

記憶手段と、前記ユーザプロフィール記憶手段に記憶されたユーザプロフィール情報と前記形態素解析手段の解析結果とに基づいて前記検索条件の多義性を解消する第2の多義性解消手段とを備え、前記第1の多義性解消手段は前記第2の多義性解消手段より優先することを特徴としている。

【0020】また、本発明に係る情報検索方法は、検索条件を入力する検索条件入力ステップと、該検索条件入力ステップで入力された検索条件から単語を抽出する形態素解析ステップと、複数の単語間の共起関係を語義と関連付けて記憶された共起情報と前記形態素解析ステップでの解析結果とに基づいて前記検索条件の共起関係を抽出し多義性を解消する第1の多義性解消ステップと、該第1の多義性解消ステップで多義性の解消された語義に基づいて情報検索を行う情報検索ステップとを含むことを特徴とし、さらにユーザの嗜好を表現するユーザプロフィール情報と前記形態素解析ステップの解析結果とに基づいて前記検索条件の多義性を解消する第2の多義性解消ステップを備え、前記第1の多義性解消ステップは前記第2の多義性解消ステップより優先することを特徴としている。

【0021】また、本発明に係る記憶媒体は、前記第1の多義性解消手順より優先される、ユーザの嗜好を表現するユーザプロフィール情報と前記係り受け解析ステップの解析結果とに基づいて前記検索条件の多義性を解消する第2の多義性解消手順が記憶されていることを特徴とし、さらに前記第2の多義性解消手順で多義性の解消された語義に対し拡張処理を行う語義拡張手順が記憶されていることを特徴としている。

【0022】尚、本発明のその他の特徴は、下記の発明の実施の形態の記載より明らかとなろう。

【0023】

【発明の実施の形態】次に、本発明の実施の形態を図面に基づいて詳説する。

【0024】図1は本発明に係る情報検索装置としての文書検索装置の一実施の形態を示すブロック構成図であって、該文書検索装置は、キーボードやマウス等からなる入力装置1と、CRTや液晶ディスプレイ等からなる表示装置2と、後述する所定のデータが格納されたハードディスク(HD)3と、フレキシブルディスク(FD)やCD(コンパクトディスク)、DVD(デジタルビデオディスク)等の外部記憶媒体にアクセスするためのリムーバブルディスクドライブ4と、通信回線を介して外部とデータ交換を行うモデムやLANコントローラ等の通信装置5と、後述する所定の制御プログラムが格納された読み出し専用の固定メモリ(ROM)6と、各種データを一時的に記憶したりワークエリアとして使用される書き込み可能なランダムアクセスメモリ(RAM)7と、バス8を介して上記各構成要素に接続され装置全体を制御する中央演算処理装置(CPU)9とを備えてい

る。

【0025】また、HD3には、語義ベースで単語の意味ベクトルが格納された単語ベクトル辞書3a、検索対象となる文書データベース3b、ユーザの嗜好を記録したユーザプロフィール3c、係り受け単語間の共起関係が格納された共起データベース3d、及びクエリの類義語展開や関連展開を行うクエリ拡張辞書3eが格納されている。

【0026】尚、本実施の形態では、CPU9で演算処理される制御プログラムは、ROM6に記憶されているが、該ROM6に代えてHD3に記憶させ、該HD3からRAM7上にロードして実行してもよく、或いはFD等の外部記憶媒体に記憶させリムーバブルディスクドライブ4を介してRAM7上にロードし実行するようにしてもよい。

【0027】図2は表示装置2の表示画面の一例を示した図であって、該表示画面はクエリ(検索条件)を表示するクエリ表示部2aと、クエリに基づいた検索結果を表示する検索結果表示部2bとから構成されている。

【0028】クエリ表示部2aは、具体的には、ユーザが入力装置1を介して入力した検索条件、例えば、自然文(「フォームの種類」「フォームの種類について知りたい」「投球フォームのバリエーション」等)、複数のキーワードの羅列(「フォーム、種類」等)、又はユーザの指示する既存文書(「文書番号267」等)が表示される。

【0029】そして、検索結果表示部2bには、検索結果として各文書を識別する文書ID、及び文書IDに対応した文書タイトル、及びクエリに対する文書IDの類似度が表示される。

【0030】図3はHD3に格納される単語ベクトル辞書3aのフォーマット図ある。

【0031】単語ベクトル辞書3aは、各単語の語義を示す意味ベクトル(意味分類ごとの特徴量リスト)の集合であって、各次元(1、2、3、…)は意味分類を表現している。

【0032】すなわち、単語ベクトル辞書3aでは、特定の語義が各単語(1、2、3、…)に意味付けられており、各単語に対して各次元の意味分類がどの程度含意されているか、つまり意味ベクトルの特徴量がマトリクス状に書き込まれている。

【0033】例えば、次元3は「宇宙・空」という意味分類を示し、次元4は「取引・売買」という意味分類を示し、次元7は「身振り・動作」という意味分類を示し、一方、単語7は「フォーム(帳票)」という特定の語義に意味付けられている。そして、該単語ベクトル辞書3aでは単語7における次元3の意味ベクトルの特徴量は「0」であるため、「フォーム(帳票)」という単語には「宇宙・空」の意味を全く有していないことが分かる。

【0034】また、単語7では次元4の特徴量が「21」と他の特徴量に比べて相対的に大きく、単語7における次元7の特徴量は「1」と相対的に小さいが、これは「フォーム（帳票）」が「取引・売買」という意味の寄与度は大きい、「身振り・動作」という意味の寄与度は小さいことを示している。

【0035】また、単語8は「フォーム（姿勢）」という語義を有しており、単語8においては次元4の特徴量は「0」であり、次元7の特徴量は「23」と相対的に大きい。これは「フォーム（姿勢）」には「取引・売買」という意味が全く存在しないが、「身振り・動作」という意味の寄与度は大きいことを示している。

【0036】このように単語ベクトル辞書3aにより、語義別の各単語の意味する寄与度を認識することができる。

【0037】図4はHD3に格納された文書データベース3bのフォーマット図であって、該文書データベース3bには文書ベクトルの特徴量が書き込まれている。

【0038】文書の意味は文書中でどのような単語が使用されたかによって決定されると判断し、各文書の意味は、その文書を構成する単語の意味ベクトルを加算していくことで算出される。したがって、算出されたベクトルの次元は単語ベクトル辞書3aの意味ベクトルの次元と同一となり、特定の意味分類を表現する。そして、加算されて得られたベクトルは「1」を基準に正規化され、該正規化されたベクトルが文書ベクトルの特徴量として文書データベース3bに格納される。

【0039】この図4から明らかなように、例えば、文書IDが「6949」の場合では次元4の特徴量は「0.009」であり、次元7の特徴量は「0.425」であり、文書IDが「6953」の場合では次元4の特徴量は「0.362」、であり、次元7の特徴量は「0.008」である。そしてこれにより文書IDが「6949」の文章は、「身振り・手振り」の意味分類は或る程度含んでいるが、「取引・売買」の意味分類を殆ど含んでおらず、また、文書IDが「6953」の文章は「取引・売買」の意味分類は或る程度含んでいるが、「身振り・動作」の意味分類をほとんど含んでいないことが分かる。

【0040】図5はHD3に格納されたユーザプロフィール3cのフォーマット図である。

【0041】ユーザプロフィール3cも単語ベクトル辞書3aの意味ベクトルと同一の次元を有し、ユーザがドキュメントファイルにアクセスする毎にプロフィールが更新される。

【0042】すなわち、初期状態ではプロフィールは「0」に設定されているが、ユーザが特定のドキュメントにアクセスすると、当該ドキュメントの文書ベクトルが算出され、算出された値が累積プロフィールに加算される。そして、新たな累積プロフィールが得られた後、

プロフィールは「1」を基準にして正規化され、正規化プロフィールの更新が行なわれる。

【0043】図5（a）は、例えばスポーツに関心のあるユーザのプロフィールを示しており、次元7（身振り、動作）の特徴量が「0.186」と比較的大きくなっている。これは該ユーザが「身振り・動作」の意味分類を有するドキュメントを多く参照していることを示している。

【0044】一方、図5（b）は窓口業務に関心の深いユーザのプロフィールを示しており、次元7（身振り、動作）の特徴量は「0.000」であるが、次元4（取引・売買）の特徴量は「0.329」と大きな数値を示している。これは該ユーザが「取引・売買」に関するドキュメントを多く参照していることを示している。

【0045】図6はHD3に格納される共起データベース3dのフォーマット図であり、係り単語、受け単語、及び両者間に介在する助詞の3つの共起情報が記憶されている。尚、前記助詞が存在しないときは「null」が書き込まれる。

【0046】本実施の形態では、入力されたクエリ中の文字列を形態素解析した後、係り受け解析を行って係り単語、受け単語、及び助詞情報を抽出し、共起データベース3dを参照し、これら係り単語、受け単語、及び助詞情報の間で照合処理を行う。そして、共起データベース3dに上記係り単語、受け単語、及び助詞情報に対応する文字列がある場合は、各単語は共起データベース3dに記載通りの語義であると解釈される。

【0047】例えば、「投球フォーム」という語句が係り受け解析によって「投球／フォーム」と抽出されたときは、このフォームの語義は共起データベース3dに従って「姿勢」の語義と解釈される。また、クエリが「フォームに情報を入力する」という場合は、係り受け解析により「フォーム／に／入力」という単語及び助詞が抽出され、これら単語及び助詞と共起データベース3dとの間で照合処理がなされ、その結果、この「フォーム」は「帳票」の語義であると解釈される。

【0048】図7はHD3に格納されるクエリ拡張辞書3eのフォーマット図であって、図7（a）は類義語辞書、図7（b）は関連語辞書を示している。

【0049】すなわち、類義語辞書には見出語に対して展開されるべき類義語が格納されている。例えば、見出語「フォーム（姿勢）」には類義語として「姿勢、形、スタイル、ポーズ」が格納され、見出語「フォーム（帳票）」には類義語として「書式、伝票、帳票」が格納されている。

【0050】尚、通常、類義語は見出語に対して同義概念または下位概念の関係にある。そして、類義語辞書はクエリ中の各単語を展開し、各単語の意味内容を許容範囲まで拡張するために使用される。従来はクエリ中に表記された「フォーム」に対し展開される類義語を有して

いたため、「姿勢」と「帳票」が混在されて展開されていたが、本実施の形態では、類義語辞書は語義ベースで保持されるので、「姿勢」と「帳票」が混在されて展開されることはない。尚、類義語にも語義情報が格納されており、類義語展開後も語義ベースで処理が可能である。

【0051】また、関連語辞書は、起点語に対して展開されるべき関連語が格納されている。例えば、起点語「フォーム（姿勢）」には関連語としては「スポーツ、分析、改善」が格納され、起点語「フォーム（帳票）」には関連語としては「購入、申し込み、振込み、送金」が格納されている。

【0052】尚、関連語は、上述した類義語とは異なり、起点語との間には上位下位の関係は存在しない。そして、関連語辞書はクエリ中の各単語を展開して、クエリ全体の意味内容がある程度充実させるために使用される。

【0053】このように構成された文書検索装置は、入力装置1からの各種の入力に応じて作動し、該入力装置1からの入力信号がCPU9に供給され、該CPU9がROM6内に記憶してある制御プログラムを読み出し、該制御プログラムに従って、各種の制御が行われる。

【0054】図8は本文書検索装置で実行される文書検索方法の処理手順の一実施の形態を示すフローチャートであって、本プログラムはCPU9で実行される。

【0055】ステップS1で各種パラメータの初期化や初期画面の点灯等、初期化処理を行った後、ステップS2では入力装置1からの操作入力を待機し、続くステップS3では入力された操作内容を判別する。

【0056】すなわち、本文書検索方法の検索手順は、文書データベースへの登録処理、ユーザプロファイルの更新処理、及びクエリに応じた検索実行処理の3つに大別され、したがって、ユーザは、検索段階に応じてこれら3つの処理のいずれかを選択して入力操作する。

【0057】そして、文書データベース3bへの登録処理が指示されたときはステップS4に進んで文書登録処理を実行し、ユーザプロファイル3cの更新処理が指示されたときはステップS5に進んでプロファイル更新処理を実行し、検索実行処理が指示されたときはステップS6に進んで検索実行処理を実行し、その後ステップS7に進んで上記の各処理の処理結果を表示パターンに展開して出力し、ステップS2に戻る。

【0058】図9はステップS4（図8）で実行される文書登録処理の処理手順を示すフローチャートであって、後述する検索処理を実行するために文書ベクトルを文書データベース3bに登録する。

【0059】ステップS11では入力されたクエリから形態素解析を行って単語の抽出処理をし、次いで係り受け解析を行う。そして続くステップS11では係り受け解析により解析された係り単語及び受け単語と共起デー

タベース3dとを照合し、当該係り単語及び受け単語を組にした文字列が共起データベース3dに格納されている場合は単語の語義を特定する。

【0060】尚、語義が特定できなかった単語についてはその表記を有する全ての語義の単語ベクトルに頻度別の重みをつけて加算される。

【0061】次に、ステップS13では文書ベクトルの生成処理を行う。すなわちステップS11とステップS12で抽出された単語及び特定された語義から単語ベクトル辞書3aを検索して意味ベクトルの特徴量を算出し、その総和から文書の特徴付ける文書ベクトルの特徴量を生成する。すなわち、文書ベクトルは、上述したように文書の表現する意味を表すものであり、各単語に関し単語ベクトル辞書3aに書き込まれた意味ベクトルの特徴量を加算していくことにより生成される。

【0062】そして続くステップS14では文書データベース3bへの登録処理を行い、メインルーチン（図8）に戻る。すなわち、文書の内容とステップS13で得られた文書ベクトルの特徴量を文書データベース3bに登録すると共に該文書データベース3bのインデックスを更新する。

【0063】図10はステップS5（図8）で実行されるプロファイル更新処理の処理手順を示すフローチャートであって、ユーザからの指示により特定のドキュメントファイルにアクセスするとき、例えば、文書データベース3bに登録されていない個人使用のFD等の外部記憶媒体へのファイルの読み書き、或いはインターネットを介したWebページにアクセスするとき等に実行される。

【0064】ステップS21では文書データを入手し、次いで、ステップS22で形態素解析により単語を抽出した後、係り受け解析を行う。次いで、ステップS23では、上述と同様、共起データベース3cを参照し、係り単語と受け単語の組が共起データベース3cに格納されている場合は単語の語義を特定する。

【0065】次に、ステップS24では、上述と同様、ステップS21とステップS22とで抽出された単語及び特定された語義から単語ベクトル辞書3aを検索して意味ベクトルを生成し、その後文書ベクトルを生成する。ステップS25では生成された文書ベクトルを累積プロファイルに加算し、続くステップS26では「1」を基準にして累積プロファイルを正規化し、これにより正規化プロファイルを作成する。

【0066】このようにしてユーザプロファイルを更新した後、ステップS27では本来の処理である各ファイルの処理（例えば、ファイルの参照処理、書き込み処理など）を行い、その後メインルーチンに戻る。

【0067】図11はステップS6（図8）で実行される検索実行処理の処理手順のフローチャートである。

【0068】ステップS31は検索クエリ入力処理を実

行し、ユーザは自然文や複数のキーワード或いは既存の文書指定等によりクエリを入力し、入力内容に応じたクエリのテキストストリングを入手する。例えば、クエリとして既存の文書を指定した場合は該文書にアクセスし、適当なフォーマットに変更して当該文書の内容をテキストファイル化し、そのテキストストリングを入手する。

【0069】次いで、ステップS32では前記テキストストリングに基づいてクエリベクトルの生成処理を行う。

【0070】図12はステップS32で実行されるクエリベクトル生成処理ルーチンのフローチャートである。

【0071】すなわち、ステップS41ではユーザ指定の検索クエリから単語を抽出し、形態素解析用辞書を使用して形態素解析を行い、更に係り受け解析を行う。続くステップS42では全ての係り受け解析の結果と共起データベース3dとを照合し、解析された係り単語と受け単語との組が共起データベース3dに格納されている場合は単語の語義を特定する。

【0072】次に、ステップS43ではクエリ中の全ての単語の多義性が解消されたか否かを判断し、解消されている場合は直ちにステップS45に進む一方、解消されていない場合はステップS44に進んでユーザプロフィールに基づく多義解消を行い、その後、ステップS45に進む。

【0073】具体的には、ステップS44では多義解消されなかったクエリ中に表記された単語の全ての語義を示す単語ベクトルXと、正規化ユーザプロフィールベクトルQとの余弦測度SD(X, Q)を求め、該余弦測度SD(X, Q)を類似度として算出する。すなわち、単語ベクトルXは、選択される語義が1つだけの場合もあり、また複数存在する場合もあり、一般的には数式

(1) 示すようにn次元(x1~xn)のベクトルで表される。同様に正規化ユーザプロフィールベクトルQも数式(2)に示すようにn次元(q1~qn)のベクトルで表され、また、余弦測度SD(X, Q)は両ベクトルの内積を両ベクトルの絶対値の積で除算した値となる。しかるに単語ベクトルX及び正規化ユーザプロフィールベクトルQは「1」を基準に正規化されているため、余弦測度SD(X, Q)は前記内積に相当し、したがって、余弦測度SD(X, Q)は、数式(3)に示すように、両ベクトルの同次元の特徴量の積の総和となる。

【0074】

【数1】

$$\vec{X} = (x1, x2, x3, \dots, xn) \quad \dots(1)$$

$$\vec{Q} = (q1, q2, q3, \dots, qn) \quad \dots(2)$$

$$\begin{aligned} SD(X, Q) &= \cos\theta \\ &= (\vec{X} \cdot \vec{Q}) / (|\vec{X}| \cdot |\vec{Q}|) \\ &= \vec{X} \cdot \vec{Q} \\ &= \sum_{k=1}^n (xk \cdot qk) \quad \dots(3) \end{aligned}$$

【0075】このようにして余弦測度SD(X, Q)、すなわち類似度を求め、ある閾値以上に類似する語義を選択して無関係と解される語義を除外することにより、ユーザプロフィールによる多義解消を行う。

【0076】このように、ステップS42で共起データベースにより多義解消できなかった単語に対してユーザプロフィールによる多義解消を行う。ここで強調すべき点は、ユーザプロフィールによる多義解消に優先して共起情報による多義解消が行われることである。

【0077】そして、ステップS45では、クエリ拡張辞書3eを使用してクエリの拡張を行う。すなわち、ユーザの指示に従い、「類義語展開のみ」、「類義語展開+関連語展開」、「関連語展開のみ」などのバリエーション処理を行う。

【0078】次に、ステップS46では検索クエリの文書ベクトル生成処理を行う。すなわち、これまでの処理で抽出された単語及び特定された語義から単語ベクトル辞書3aを検索し、単語ごとの次元別の特徴量を算出し、その総和から文書ベクトルを生成して図11のルーチンに戻る。

【0079】次に、図11のステップS33ではステップS32で得られたクエリベクトルとQ'と検索対象となる文書データベース3bの文書ベクトルX'とから余弦測度SD(X', Q')、すなわち類似度を算出し、RAM7に格納する。

【0080】図13はステップS33で実行される類似度生成処理ルーチンのフローチャートである。

【0081】すなわち、ステップS51では文書データベース3b内の検索対象となる文書を指定するカウンタのカウント値Nを初期値1にセットし、続くステップS52では文書データベース3bを検索し、N番目(最初のループではN=1)の文書の文書ベクトルX'を読み出し、ステップS53ではN番目の文書ベクトルX'と検索クエリのクエリベクトルQ'とに基づいて類似度を算出する。すなわち、ステップS53では検索クエリを1つの文書とみなしてクエリベクトルQ'を求め、検索対象の文書データベース3b上の文書の文書ベクトルX'とクエリベクトルQ'との余弦測度SD(X', Q')を求め、類似度を算出する(下記数式(4)~(6)参照)。

【0082】

【数2】

$$\bar{X}' = (x'1, x'2, x'3, \dots, x'n) \quad \dots(4)$$

$$\bar{Q}' = (q'1, q'2, q'3, \dots, q'n) \quad \dots(5)$$

$$\begin{aligned} SD(X', Q') &= \cos \theta' \\ &= (\bar{X}' \cdot \bar{Q}') / (|\bar{X}'| \cdot |\bar{Q}'|) \\ &= \bar{X}' \cdot \bar{Q}' \\ &= \sum_{k=1}^n (x'k \cdot q'k) \quad \dots(6) \end{aligned}$$

【0083】そして、例えば、「フォーム」という表記を上記図5(a)(b)の2つのプロファイルに従って解釈した場合、図5(a)のプロファイルに従うと、次元7の特徴量が大きいので、図3の単語7と単語8とは、次元7の特徴量が大きい単語8(「フォーム(姿勢)」)との内積が大きくなり、「姿勢」の語義が採用される。

【0084】一方、図5(b)のプロファイルの従うと、次元4の特徴量が大きいので、図3の単語7と単語8とは、次元4の特徴量が大きい単語7「フォーム(帳票)」との内積が大きくなり、「帳票」の語義が採用される。

【0085】次に、ステップS54では算出された類似度をRAM7に格納し、続くステップS55では文書データベース3b内の検索対象文書に残文書があるか否かを判断し、存在しない場合はそのまま図11のルーチンに戻る一方、存在する場合はステップS56でカウンタのカウンタ値Nを「1」だけインクリメントしてステップS52に戻り、上述の処理を繰り返す。

【0086】尚、本実施の形態では前記類似度をRAM7に格納しているが、文書データベース3bに登録してもよい。

【0087】次に、図11のルーチンに戻り、ステップS34ではRAM7を参照し、ステップS33で得られた文書ごとの類似度を順序付けする。そして、ステップS35ではステップS34で順序付けされた文書を検索結果としてリストアップし、表示装置2に表示する。尚、この時、ステップS33で登録された類似度の値も同時に表示する。

【0088】このように本実施の形態によれば、共起データベース3dに従ってクエリを解析することにより、クエリの語義を正確に解釈することができ、これによりクエリの多義性解消を高精度に行なうことができ、また、共起データベース3dに基づいた多義解消を行なうことができない場合は、ユーザプロフィール情報によって、ユーザの過去の嗜好に従って多義解消することができるので、よりユーザの意図に即した検索精度の高い検索結果を容易且つ迅速に得ることができる。

【0089】図14は本発明に係る情報検索装置としての文書検索装置の第2の実施の形態を示すブロック構成図であって、本第2の実施の形態ではクエリの拡張を行わない場合を示しており、クエリ拡張辞書が省略されて

いる。

【0090】図15は本第2の実施の形態におけるクエリベクトル生成処理ルーチンのフローチャートであって、ステップS61ではユーザ指定の検索クエリから単語を抽出し、形態素解析用辞書を使用して形態素解析を行い、更に係り受け解析を行う。続くステップS62では全ての係り受け解析の結果と共起データベース3dとを対照し、解析された係り単語と受け単語との組が共起データベース3dに格納されているかどうかに応じて各語義のもつともらしさ(語義尤度)を算出していく。

【0091】次に、ステップS63では上記第1の実施の形態と同様、余弦測度に基づく類似度を算出し、類似度に従って語義尤度を求めていく。そしてステップS62、S63で求めた語義尤度を合算し、最終的に最ももつともらしいとされた語義を選択する。このとき、語義尤度の重みとしてS62の語義尤度に対する重みをより大きくすることで、共起情報による多義解消をプロフィールによる多義解消よりも優先する。

【0092】そして、ステップS64では、検索クエリの文書ベクトル生成処理を行う。すなわち、これまでの処理で抽出された単語及び特定された語義から単語ベクトル辞書3aを検索し、単語ごとの次元別の特徴量を算出し、その総和から文書ベクトルを生成している。

【0093】このように本実施の形態は、クエリ拡張を行わない場合の例を示すと共に、共起データベースに基づく多義解消を、ユーザプロフィール情報による多義解消よりも優先する別の実施形態を示している。

【0094】尚、本発明は上記実施の形態に限定されるのではない。

【0095】上記実施の形態では、文書検索方式として類似文書検索について説明したが、他の検索方式に適用することもできる。例えば、従来の全文検索システムにおいてもクエリを類義語展開することがあるが、本発明によればクエリが語義ベースで解析されているので、図7に示すような語義ベースの類義語辞書により容易に必要な類義語だけを展開することができ、また展開単語は特定語義についての類義語に限定されるので、検索ノイズを減少させることができる。

【0096】また、上記実施の形態では、ユーザプロフィールの更新をファイルアクセス毎に更新するように構成したが、アクセスしたファイルの履歴情報のみを記録しておき、ある所定期間ごとに一括してプロフィールを更新するように構成するようにするのも好ましい。この場合はファイルアクセス毎に余分な処理時間を要することなく快適にファイルアクセスすることができ、また、通常コンピュータを使用しない深夜などに一括してプロフィールを更新することもでき、この場合はプロフィール更新による通常ユーザ層に対する処理時間的影響はより軽微となる。

【0097】また、ユーザプロフィールをアクセスした

ファイル履歴から構成するようにしているが、ユーザが直接このユーザプロフィールを作成するようにしてもよい。

【0098】また、ユーザが直接ユーザプロフィール情報を最初から作成するのは困難な場合があるため、ある種のガイドラインを作成し、システムの質問に応じていくだけでプロフィールが作成されるようにしてもよい。

【0099】さらに、上記実施の形態では、ユーザプロフィールは各個人が別々の情報を保有するように構成したが、会社組織などにおいては各グループ単位でプロフィールを保有することも考えられ、斯かる場合はグループ内の他のメンバーによるファイルアクセスによってプロフィール情報が更新されることとなる。

【0100】上述の実施形態においては、語義の多義性を解消する方法として、共起データベースによる方法、ユーザプロフィールによる方法の2種類を挙げたが、多義解消手段はこの2種類に限定されるものではない。例えば、各単語の語義リストを表示し、該語義リストの中から所望の語義をユーザが選択するようにしてもよく、文脈に応じて語義を解釈するようにしてもよい。すなわち、ユーザの発行する数多くのクエリに対して、文脈ベクトルを用意し、クエリ入力ごとに文脈ベクトルを更新することにより、新たなクエリ入力では重要単語を省略してもユーザの意図通りの検索を行なうことが可能となり、斯かる多義解消方法を混在するようにしてもよい。

【0101】

【発明の効果】以上詳述したように本発明によれば、ベクトル空間モデルを応用した類似文書検索において共起情報に従って検索クエリを解析し多義解消するので、クエリが正確に解釈でき、ユーザの意図に沿った検索精度の高い情報検索を行なうことができる。

【0102】また、検索クエリ中の多義語をユーザの過去の嗜好を表現するユーザプロフィール情報を参照して多義解消するので、クエリをよりユーザの嗜好に合わせて解釈でき、ユーザの意図に沿った検索精度の高い情報

検索を行うことができる。また、共起情報による多義解消をユーザプロフィール情報による多義解消よりも優先するので、よりユーザの意図に沿った検索精度の高い情報検索を行なうことができる。

【図面の簡単な説明】

【図1】本発明に係る情報検索装置としての文書検索装置の一実施の形態（第1の実施の形態）を示すブロック構成図である。

【図2】表示装置の表示画面の一例を示す図である。

【図3】単語ベクトル辞書のフォーマット図である。

【図4】文書データベースのフォーマット図である。

【図5】ユーザプロフィールのフォーマット図である。

【図6】共起データベースのフォーマット図である。

【図7】クエリ拡張辞書のフォーマット図である。

【図8】本発明に係る情報検索方法としての文書検索方法の検索手順を示すメインルーチンのフローチャートである

【図9】文書登録処理ルーチンのフローチャートである。

【図10】プロフィール更新処理ルーチンのフローチャートである。

【図11】検索実行処理ルーチンのフローチャートである。

【図12】クエリベクトル生成処理ルーチンのフローチャートである。

【図13】類似度生成処理ルーチンのフローチャートである。

【図14】本発明に係る情報検索装置としての文書検索装置の第2の実施の形態を示すブロック構成図である。

【図15】第2の実施の形態におけるクエリベクトル生成処理ルーチンのフローチャートである。

【符号の説明】

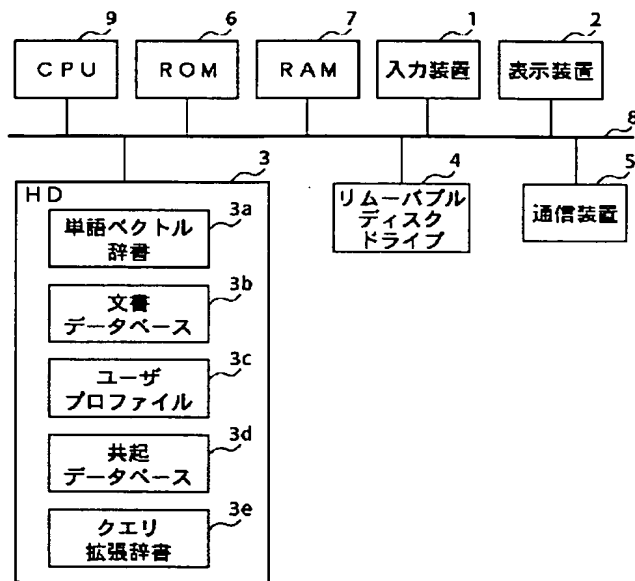
- 1 入力装置
- 3 HD
- 9 CPU

【図6】

3d

係り単語	助詞	受け単語
投球	null	フォーム (姿勢)
フォーム (帳票)	に	入力
登録	null	フォーム (帳票)

【図 1】



【図 3】

(宇宙・空) (取引・売買) (身振り・動作)

単語 \ 次元	1	2	3	4	5	6	7	8	9	10
1	0	23	32	0	0	10	0	0	0	3
2	34	0	0	0	2	0	22	21	0	0
3	4	22	0	4	0	0	0	0	9	0
4	3	12	32	4	0	7	0	0	32	0
5	0	0	0	0	0	0	0	14	0	0
6	4	0	6	8	0	0	0	0	0	0
7 (フォーム(帳票))	4	7	0	21	0	0	1	3	5	0
8 (フォーム(姿勢))	1	5	0	0	0	3	23	0	8	0
9	1	0	0	0	0	0	12	0	0	0
10	0	0	0	21	0	0	14	0	3	0
11	0	10	0	9	0	0	21	0	0	0
12	12	0	34	6	0	12	14	24	32	0
13	0	0	19	0	0	0	22	32	0	6
14	0	0	0	6	0	0	5	0	0	0
15	13	11	34	3	8	0	0	44	0	12
....											

【図 2】

2a

文書ID	文書タイトル	類似度
1028	タイトル1028	95
877	タイトル877	93
23988	タイトル23988	90
10029	タイトル1002	83
2897	タイトル2897	82
9999	タイトル9999	82
88	タイトル88	80
11987	タイトル11987	77
14576	タイトル14576	74
18726	タイトル18726	74
29837	タイトル29837	73
23874	タイトル23874	73
10098	タイトル10098	70
8729	タイトル8729	70
9872	タイトル9872	70
5456	タイトル5456	65

2b

【図 7】

(a)

見出語	類義語
フォーム(姿勢)	姿勢、形、スタイル、ポーズ
フォーム(帳票)	書式、伝票、帳票
椅子	腰掛け、チェア、ソファ、ベンチ
⋮	⋮

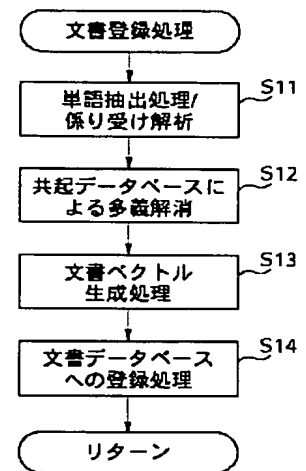
(b)

起点語	関連語
フォーム(姿勢)	スポーツ、分析、改善
フォーム(帳票)	購入、申し込み、振り込み、送金
⋮	⋮

【図4】

		次元										
		1	2	3	4	5	6	7	8	9	10
文書ID	6947	0.183	0.214	0.000	0.006	0.635	0.000	0.021	0.009	0.021	0.014
	6948	0.035	0.025	0.000	0.009	0.301	0.115	0.029	0.005	0.128	0.019
	6949	0.035	0.025	0.000	0.009	0.301	0.115	0.425	0.029	0.128	0.019
	6950	0.000	0.000	0.000	0.000	0.496	0.369	0.050	0.003	0.165	0.000
	6951	0.110	0.154	0.000	0.007	0.724	0.086	0.000	0.000	0.000	0.000
	6952	0.142	0.087	0.040	0.008	0.577	0.428	0.098	0.003	0.094	0.044
	6953	0.095	0.055	0.087	0.362	0.532	0.223	0.008	0.062	0.059	0.024
	6954	0.309	0.087	0.000	0.009	0.294	0.119	0.000	0.009	0.070	0.000
	6955	0.411	0.087	0.000	0.002	0.482	0.027	0.000	0.004	0.083	0.116
	6956	0.073	0.036	0.000	0.005	0.514	0.247	0.000	0.007	0.000	0.000
	6957	0.014	0.087	0.027	0.002	0.038	0.152	0.027	0.002	0.100	0.014
	6958	0.130	0.308	0.036	0.005	0.051	0.395	0.036	0.005	0.036	0.018
	6959	0.169	0.038	0.000	0.002	0.169	0.103	0.000	0.004	0.131	0.120
											

【図9】



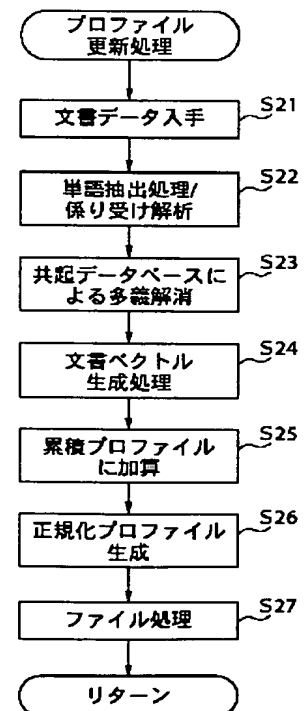
【図5】

		次元										
		1	2	3	4	5	6	7	8	9	10
累積		1205	551	3338	42	65	366	3282	2318	844	80
正規化		0.068	0.031	0.189	0.002	0.003	0.024	0.186	0.131	0.048	0.004

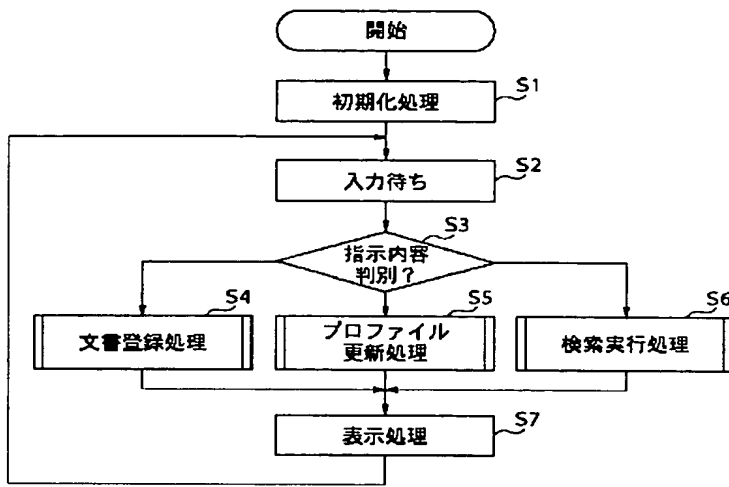
(b)

		次元										
		1	2	3	4	5	6	7	8	9	10
累積		35	13	36	5835	62	1456	8	3	23	367
正規化		0.002	0.003	0.002	0.329	0.003	0.082	0.000	0.000	0.001	0.021

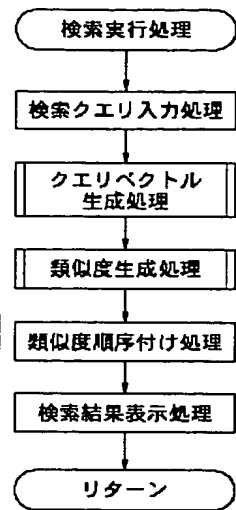
【図10】



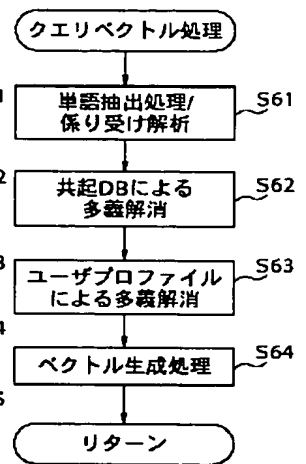
【図8】



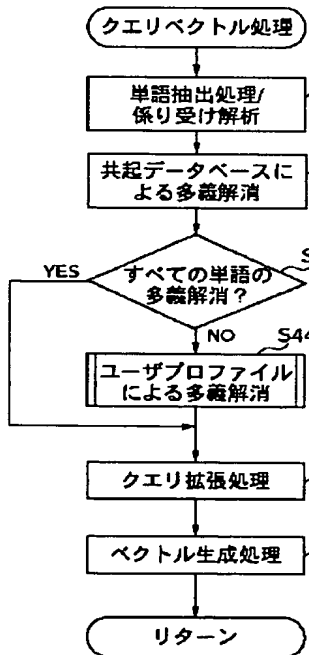
【図11】



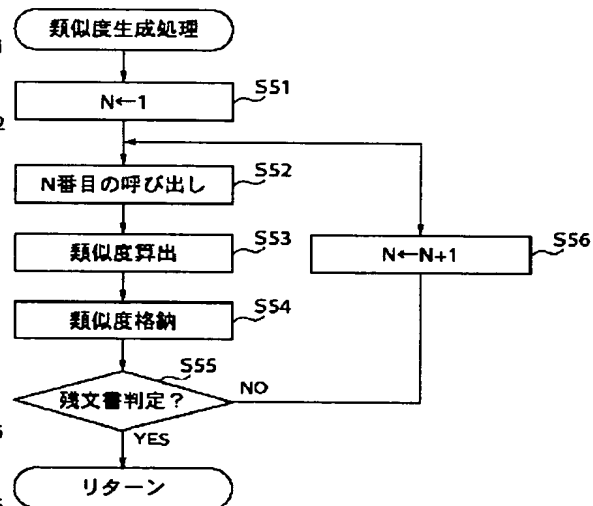
【図15】



【図12】



【図13】



【図 14】

